# Spectra of large block matrices

R. Rashidi Far      T. Oraby      W. Bryc[*]      R. Speicher[‡]

## Abstract

In a frequency selective slow-fading channel in a MIMO system, the channel matrix is of the form of a block matrix. This paper proposes a method to calculate the limit of the eigenvalue distribution of block matrices if the size of the blocks tends to infinity. While it considers random matrices, it takes an operator-valued free probability approach to achieve this goal. Using this method, one derives a system of equations, which can be solved numerically to compute the desired eigenvalue distribution.

The paper initially tackles the problem for square block matrices, then extends the solution to rectangular block matrices. Finally, it deals with Wishart type block matrices. For two special cases, the results of our approach are compared with results from simulations. The first scenario investigates the limit eigenvalue distribution of block Toeplitz matrices. The second scenario deals with the distribution of Wishart type block matrices for a frequency selective slow-fading channel in a MIMO system for two different cases of $n_R = n_T$ and $n_R = 2n_T$. Using this method, one may calculate the capacity and the Signal-to-Interference-and-Noise Ratio in large MIMO systems.

**Keywords:** MIMO systems, channel models, eigenvalue distribution, fading channels, free probability, Cauchy transform, intersymbol interference, random matrices, Stieltjes transform.

# 1   Introduction

With the introduction of some sophisticated communication techniques such as CDMA (Code-Division Multiple-Access) and MIMO (Multiple-Input Multiple-Output), the communications community has been looking into analyzing different aspects of these systems, ranging from the channel capacity to the structure of the receiver. It has been shown that the channel matrix plays a key role in the capacity of the channel [1, 2] as well as in the structure of the optimum receiver [3, 4]. More precisely, the eigenvalue distribution of the channel matrix is the factor of interest in different applications.

For a MIMO wireless system with $n_T$ transmitter antenna and $n_R$ receiver antenna, the received signal at time index $n$, $Y_n = [y_{1,n}, \cdots, y_{n_R,n}]^T$, will be as follows:

$$Y_n = HX_n + N_n, \tag{1}$$

where $H$ is the channel matrix, $X_n = [x_{1,n}, \cdots, x_{n_T,n}]^T$ is the transmitted signal at time $n$ and $N_n$ is the noise signal. The channel matrix entries $h_{i,j}$'s reflect the channel effect on the signal transmitted from antenna $j$ in the transmitter and received at antenna $i$ in the receiver. In a more realistic channel modeling, one may consider the Intersymbol-Interference (ISI) [5, Chapter 2]. In this case, the channel impulse response between the transmitter antenna $j$ and the receiver antenna $i$ is a vector $h_{ij} = \begin{bmatrix} h_1^{(ij)} & h_2^{(ij)} & \cdots & h_{L-1}^{(ij)} & h_L^{(ij)} \end{bmatrix}^T$ where $L$ is the length of the impulse response of the channel (number of the taps). Consequently, the channel matrix for a signal frame of $K$ will be as follows:

$$H = \begin{bmatrix} A_1 & A_2 & \cdots & A_L & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_1 & A_2 & \cdots & A_L & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_1 & A_2 & \cdots & A_L & \mathbf{0} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \cdots & & \ddots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{0} & A_1 & A_2 & \cdots & A_L \end{bmatrix}, \tag{2}$$

where there are $K-1$ zero-matrices in each row and $A_l = (h_l^{(ij)})_{\substack{i=1,\cdots,n_R \\ j=1,\cdots,n_T}}$ (see Fig. 1 for the block diagram). To calculate the capacity of such a channel, one needs to know the eigenvalue distribution of the $H^*H$ [6].

Free probability and random matrix theory have proven to be reliable tools in tackling similar problems [6]. Tse and Zeitouni [7] applied random

2

matrix theory to study linear multiuser receivers, Moustakas *et. al.* [8] applied it to calculate the capacity of a MIMO channel. Müller [9] employed it in calculating the eigenvalue distribution of a particular fading channel and later Debbah and Müller [10] applied it in MIMO channel modeling.

While the most basic random matrices are Gaussian and Wishart random matrices, and it is now quite widely known (see, for example, [6]) how to use tools from free probability (like the $R$- or $S$-transform [11]) to deal with such random matrices, there are many interesting situations, as the matrix $H$ above, which cannot be modeled with these simple kind of random matrices. Very often these can, at least in first approximation, be assumed to be block random matrices. These are matrices built out of blocks, where each block is a Gaussian random matrix; however, the variance of the entries might be different from block to block and, in particular, some of the blocks might repeat at different places in the matrix. Our main example for such block matrices is the channel matrix $H$ as in (2), but there are also other interesting matrices of this kind. In particular, we will also consider block Toeplitz matrices, like the $3 \times 3$ block matrix

$$X = \begin{bmatrix} A & B & C \\ B & A & B \\ C & B & A \end{bmatrix},$$

where $A, B, C$ are independent selfadjoint (which we use as a synonym of Hermitian) $N \times N$ Gaussian random matrices. Such matrices were considered, e.g., in [12], but their limiting eigenvalue distribution for $N \to \infty$ remained open.

While Müller[13] has addressed a similar problem using free probability, in this paper we will show how to use a more general version of free probability theory, so-called operator-valued free probability theory, to calculate the asymptotic eigenvalue distribution of such block matrices. It was the fundamental insight of Shlyakhtenko [14, 15] that operator-valued free probability can be used for dealing with special block matrices. Our main theorems could, by building on Shlyakhtenko's work, be derived quite directly by combining various results from the literature on operator-valued free probability theory. However, since most readers will not be familiar with operator-valued free probability theory, we prefer to give also a more direct proof which does not assume any prior knowledge on operator-valued free probability. Actually, the "operator-valued" structure of the result will arise quite canonically in our proof, and thus our presentation will hopefully convince the reader not

3

only of the power but also of the conceptual beauty of operator-valued free probability theory and serve as an appetizer to read more about this field.

In Section 2 some basic notations and definitions that are used through the paper are introduced [16]. The theorems are stated in Section 3 while they are proved in Section 4. In Section 5 it is shown as a warm up exercise how the Marchenko-Pastur law can be derived using the proposed method. Then, two general scenarios are studied in this section. In the first scenario the eigenvalue distribution of block Toeplitz matrices is computed and numerical results are shown for $3 \times 3$, $4 \times 4$ and $5 \times 5$ block Toeplitz matrices. The second example computes the eigenvalue distribution for a MIMO system with Intersymbol-Interference (ISI) in two different cases. Section 5 concludes with a discussion on the convergence and numerical behaviour of the proposed method. Conclusion remarks are drawn in Section 6.

# 2 Preliminaries

## 2.1 Notations

The following notations are adopted in the paper:

$\otimes$      Tensor (Kronecker) product of matrices
$\oplus$      direct sum
$\Im(X)$      Imaginary part of $X$
$I_d$      $d \times d$ Identity matrix
$\overline{X}$      complex conjugate of $X$
$\delta_{ij}$      Dirac delta function
$X^*$      Hermitian conjugate of matrix $X$

## 2.2 Gaussian family and Wick formula

The entries of our random matrices will consist of Gaussian random variables; since entries might repeat at various places in the matrix, not all entries are independent and we use the notion of a Gaussian family to describe this general situation. A Gaussian family is a family of random variables $x_1, \ldots, x_n$ which are linear combinations of independent Gaussian random variables. Clearly, the whole information about the joint distribution of such a family is contained in the covariance matrix $C = (c_{ij})_{i,j=1}^{n}$, $E[x_i x_j] = c_{ij}$. There is a very precise combinatorial formula, usually called Wick formula,

4

which allows to express the higher joint moments in a Gaussian family in terms of second moments (i.e., in terms of the covariance matrix). Namely, for all $k \in \mathbb{N}$ and $1 \leq i(1), \cdots, i(k) \leq n$, we have:

$$E\left[x_{i(1)} \cdots x_{i(k)}\right] = \sum_{\pi \in \mathcal{P}_2(k)} \prod_{(r,s) \in \pi} E\left[x_{i(r)} x_{i(s)}\right], \qquad (3)$$

where $\mathcal{P}_2(k)$ is the set of all pairings of the set $\{1, \cdots, k\}$. Since there is no pairing of an odd number of elements, this formula also contains the statement that all odd moments of a Gaussian family vanish.

## 2.3  Cauchy transform and Stieltjes inversion formula

Our main analytical object for determining the eigenvalue distribution of a random matrix is the Cauchy transform $G_\mu$ of a probability measure $\mu$ on $\mathbb{R}$. It is defined as follows:

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-t} d\mu(z), \quad z \in \mathbb{C}^+, \qquad (4)$$

where $\mathbb{C}^+ := \{s + it \,|\, s, t \in \mathbb{R}, t > 0\}$. $G_\mu$ is analytic and takes values in $\mathbb{C}^- := \{s + it \,|\, s, t \in \mathbb{R}, t < 0\}$. For a compactly supported $\mu$, $G_\mu$ can be expanded as a power series:

$$G_\mu(z) = \sum_{n=0}^{\infty} \frac{\alpha_n}{z^{n+1}}, \quad |z| > r, \qquad (5)$$

where $r := \sup \{|t| \mid t \in \mathrm{supp}(\mu)\}$ and $\alpha_n := \int_{\mathbb{R}} t^n d\mu(t)$ is the $n$th moment of $\mu$ for $n \geq 0$. Using (4), it is clear that the Cauchy transform has the following property:

$$\lim_{z \in \mathbb{C}^+, |z| \to \infty} z G_\mu(z) = 1. \qquad (6)$$

One can recover the probability measure $\mu$ from its Cauchy transform by the following formula, known as the Stieltjes inversion formula:

$$d\mu(t) = \lim_{\epsilon \to 0} h_\epsilon(t) \, dt$$

where

$$h_\epsilon(t) := -\frac{1}{\pi} \Im G_\mu(t + i\epsilon), \quad \forall t \in \mathbb{R}. \qquad (7)$$

# 3 Statement of the Theorems

## 3.1 Selfadjoint block matrix with square blocks

We will consider $d \times d$-matrices with a given block structure; the blocks themselves will be large Gaussian $N \times N$-random matrices; our aim is to calculate the limit eigenvalue distribution for such block matrices if the size $N$ of the Gaussian matrices tends to infinity. The main point is that the blocks are not all different, but they might repeat (as themselves or their adjoint) at different places. Otherwise from that they should be independent.

We write our matrices as

$$X_N = \sum_{i,j=1}^{d} E_{ij} \otimes A^{(i,j)}, \tag{8}$$

where $E_{ij}$ are the elementary $d \times d$-matrices having a 1 at the entry $(i,j)$ and 0 otherwise. For each $i, j = 1, \ldots, d$ we have a Gaussian $N \times N$ random matrix $A^{(i,j)}$, which we call a block. In each block $A^{(i,j)}$ all the entries are Gaussian and independent. Furthermore, we want $X_N$ to be selfadjoint, which means that $A^{(i,j)} = A^{(j,i)*}$ for all $i, j = 1, d$. In particular, the blocks $A^{(i,i)}$ on the diagonal must be selfadjoint. We would like to leave it open whether the blocks $A^{(i,j)}$ with $i \neq j$ are selfadjoint or not - this depends on the concrete situation.

A convenient way to encode the information about which blocks agree, is by specifying the covariance $\sigma(i, j; k, l)$ between an entry in the matrix $A^{(i,j)}$ and an entry in the matrix $A^{(k,l)}$. In the case that the two blocks are either independent or the same (where we also include the possibility that one is the adjoint of the other), the covariance $\sigma$ takes just on the values 0 or 1. However, we can also be more general and allow arbitrary $\sigma$'s. The only constraint comes from the requirement that $X_N$ is selfadjoint which means that we must have

$$\sigma(i, j; k, l) = \overline{\sigma(k, l; i, j)} \tag{9}$$

for all $i, j, k, l = 1, \ldots, d$.

If we denote the entries of the matrix $A^{(i,j)}$ by $a_{rp}^{(i,j)}$ with $r, p = 1, \ldots, N$, then we can summarize all the above by requiring that the collection of all entries $\{a_{rp}^{(i,j)} \mid i, j = 1, \ldots, d, \ r, p = 1, \ldots, N\}$ of the matrix $X_N$ forms a

Gaussian family which is determined by

$$a_{rp}^{(i,j)} = \overline{a_{pr}^{(j,i)}} \qquad \text{for all } i, j = 1, \ldots, d, \; r, p = 1, \ldots, N$$

and the prescription of mean zero and covariance

$$E[a_{rp}^{(i,j)} a_{qs}^{(k,l)}] = \frac{1}{n} \delta_{rs} \delta_{pq} \cdot \sigma(i, j; k, l), \tag{10}$$

where

$$n = dN.$$

Note that the above gives also the covariance between $a$'s and $\bar{a}$'s as

$$E[a_{rp}^{(i,j)} \overline{a_{sq}^{(l,k)}}] = \frac{1}{n} \delta_{rs} \delta_{pq} \cdot \sigma(i, j; k, l)$$

and thus the joint distribution of the matrix entries of $A$ is uniquely determined.

Furthermore, with this description we have the possibility of including both cases where non diagonal blocks are selfadjoint or not (or mixtures of this). Consider a situation where $\sigma$ takes on only the values 0 or 1. Then a non-diagonal $A^{(i,j)}$ is selfadjoint if and only if $\sigma(i, j; i, j) = 1$. As an example, contrast the situation

$$\begin{bmatrix} 0 & A_1 & 0 \\ A_1 & 0 & A_1 \\ 0 & A_1 & 0 \end{bmatrix}$$

where $A_1$ is a selfadjoint Gaussian $N \times N$-matrix with the situation

$$\begin{bmatrix} 0 & B_1 & 0 \\ B_1^* & 0 & B_1 \\ 0 & B_1^* & 0 \end{bmatrix},$$

where $B_1$ is a non-selfadjoint Gaussian $N \times N$-matrix.

In the first case $\sigma(i, j; k, l) = 1$ for all 16 possible combinations of $(i, j)$ and $(k, l)$ from the set $\{(1, 2), (2, 1), (2, 3), (3, 1)\}$, whereas in the second case $\sigma(i, j; k, l)$ is only non-vanishing for

$$\sigma(1, 2; 2, 1), \sigma(1, 2; 3, 2), \sigma(2, 3; 2, 1), \sigma(2, 3; 3, 2)$$

and the "adjoint values"

$$\sigma(2, 1; 1, 2), \sigma(3, 2; 1, 2), \sigma(2, 1; 2, 3), \sigma(3, 2; 2, 3).$$

**Notation 1.** Fix a natural number $d$ and a covariance function

$$\sigma = \big(\sigma(i,j;k,l)\big)_{i,j,k,l=1}^{d}$$

such that (9) holds.

1) $M_d(\mathbb{C})$ are just the $d \times d$-matrices with complex entries,

$$M_d(\mathbb{C}) := \{(d_{ij})_{i,j=1}^{d} \mid i,j = 1,\ldots,d\}.$$

2) We define a *covariance mapping*

$$\eta : M_d(\mathbb{C}) \to M_d(\mathbb{C})$$

as follows:
For $D = (d_{ij})_{i,j=1}^{d} \in M_d(\mathbb{C})$ we have $\eta(D) = (\eta(D)_{ij})_{i,j=1}^{d}$ with

$$[\eta(D)]_{ij} := \frac{1}{d} \sum_{k,l=1}^{d} \sigma(i,k;l,j) d_{kl}.$$

3) Furthermore, $\mathrm{tr}_d$ denotes the normalized trace on $M_d(\mathbb{C})$, i.e.,

$$\mathrm{tr}_d(D) := \frac{1}{d} \sum_{i=1}^{d} [D]_{ii}.$$

**Theorem 2.** *With the above notation, for each $N \in \mathbb{N}$, consider block matrices* (8), *where, for each $i,j = 1,\ldots,d$, the blocks $A^{(i,j)} = \big(a_{rp}^{(i,j)}\big)_{r,p=1}^{N}$ are Gaussian $N \times N$ random matrices such that the collection of all entries $\{a_{rp}^{(i,j)} \mid i,j = 1,\ldots,d,\ r,p = 1,\ldots,N\}$ of the matrix $X_N$ forms a Gaussian family which is determined by*

$$a_{rp}^{(i,j)} = \overline{a_{pr}^{(j,i)}} \qquad \text{for all } i,j = 1,\ldots,d,\ r,p = 1,\ldots,N$$

*and the prescription of mean zero and covariance* (10).

*Then, for $N \to \infty$, the $n \times n$ matrix $X_N$ has a limiting eigenvalue distribution whose Cauchy transform $G(z)$ is determined by*

$$G(z) = \mathrm{tr}_d(\mathcal{G}(z)),$$

*where $\mathcal{G}(z)$ is an $M_d(\mathbb{C})$-valued analytic function on the upper complex half plane, which is uniquely determined by the facts that*

$$\lim_{|z|\to\infty,\Im(z)>0} z\mathcal{G}(z) = I_d, \tag{11}$$

8

*and that it satisfies for all z in the upper complex half plane the matrix equation*

$$z\mathcal{G}(z) = I_d + \eta(\mathcal{G}(z)) \cdot \mathcal{G}(z). \tag{12}$$

The proof of this theorem is just a few lines if one cites the relevant literature from operator-valued free probability theory. We will indicate this below. However, since most readers will not be familiar with operator-valued free probability theory, we prefer to give in the next section a more direct proof from scratch. As one will see, the "operator-valued" structure of the result will arise quite canonically and the next section is also intended to introduce the reader to some relevant notions and ideas of operator-valued free probability theory.

Let us now give, for the reader familiar with operator-value free probability, the condensed version of the proof of our Theorem 2. First, one has to observe that the joint moments of the blocks converges to a semi-circular family, thus the wanted limit distribution of $X_N$ is the same as the one of a $d \times d$-matrix $S$, where the entries of $S$ are from a semi-circular family, with covariance $\sigma$. By using the description of operator-valued cumulants of this matrix in terms of the cumulants of the entries of the matrix (see [17] ), it is obvious that $S$ is a $M_d(\mathbb{C})$-valued semi-circular element, with covariance $\eta$. The equation for $\mathcal{G}(z)$ follows then from the basic $R$-transform or cumulant theory of operator-valued free probability theory, see [18, 19].

In the special case that blocks of our block matrix $X_N$ do not repeat (apart from the symmetry condition $X_N = X_N^*$) one might call the block matrix a band matrix (in such a situation one often lets $d$ also go to infinity). The fundamental observation that limits of Gaussian band matrices are operator-valued semicircular elements – and thus that operator-valued free probability can be used for determining their eigenvalue distribution – was made by Shlyakhtenko in [14]. There he proved actually the special case of the above Theorem 2 for band matrices. In this case, the covariance function $\eta$ maps diagonal matrices to diagonal ones and one can consider the block matrix as a semi-circular element over the diagonal matrices.

## 3.2   Selfadjoint block matrix with rectangular blocks

In many applications one is also interested in situations where the blocks themselves might not be square matrices, but more general rectangular matrices. Of course, the sizes of the blocks must fit together to make up a big

square matrix. This means that we replace $n = dN$ by a decomposition $n = N_1 + \cdots + N_d$, and the block $A^{(i,j)}$ will then be a $N_i \times N_j$-matrix. We are interested in the limit that $N_i/n$ converges to some number $\alpha_i$. It turns out that one can modify our Theorem 2 (and also its proof) quite easily from the square to the general rectangular situation. Let us first introduce the generalizations of our relevant notations from the square case. Note that dependent rectangular blocks can be re-cut into different nonequivalent configurations of dependent blocks. We will assume that such repartitioning has already been done and resulted in the covariance function $\sigma(i, j; k, l)$ that can only be different from zero if the size of the block $A^{(i,j)}$ fits (at least in the limit $n \to \infty$) with the size of the block $A^{(k,l)}$.

**Notation 3.** Fix a natural number $d$ and a $d$-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $0 < \alpha_i < 1$ for all $i = 1, \ldots, d$ and $\alpha_1 + \cdots + \alpha_d = 1$. Furthermore, let a covariance function $\sigma = \big(\sigma(i, j; k, l)\big)_{i,j,k,l=1}^d$ be given with the property that (9) holds and in addition $\sigma(i, j; k, l) = 0$ unless $\alpha_i = \alpha_l$ and $\alpha_j = \alpha_k$. Then we use the following notations.

1) $M_\alpha(\mathbb{C})$ are those matrices from $M_d(\mathbb{C})$ which correspond to square blocks,
$$M_\alpha(\mathbb{C}) := \{(d_{ij}) \in M_d(\mathbb{C}) \mid d_{ij} = 0 \text{ unless } \alpha_i = \alpha_j\}.$$

2) We define the *weighted covariance mapping*
$$\eta_\alpha : M_\alpha(\mathbb{C}) \to M_\alpha(\mathbb{C})$$
as follows:
For $D = (d_{ij})_{i,j=1}^d \in M_\alpha(\mathbb{C})$ we have $\eta_\alpha(D) = (\eta_\alpha(D)_{ij})_{i,j=1}^d$ with
$$[\eta_\alpha(D)]_{ij} := \sum_{k,l=1}^d \sigma(i, k; l, j)\alpha_k d_{kl}.$$

3) Furthermore, the weighted trace
$$\mathrm{tr}_\alpha : M_\alpha(\mathbb{C}) \to M_\alpha(\mathbb{C})$$
is given by
$$\mathrm{tr}_\alpha\big((d_{ij})_{i,j=1}^d\big) := \sum_{i=1}^d \alpha_i d_{ii}.$$

**Theorem 4.** *With the above notation, for $\{N_1, \ldots, N_d\} \subset \mathbb{N}$ consider block matrices*

$$X_{N_1,\ldots,N_d} = \sum_{i,j=1}^{d} E_{ij} \otimes A^{(i,j)}.$$

*Here $E_{ij}$ are the elementary $d \times d$-matrices having a 1 at the entry $(i,j)$ and 0 otherwise and, for each $i,j = 1, \ldots, d$, the $A^{(i,j)}$ are Gaussian $N_i \times N_j$ random matrices, $A^{(i,j)} = \left(a_{rp}^{(i,j)}\right)_{\substack{r=1,\ldots,N_i \\ p=1,\ldots,N_j}}$. The latter are such that the collection of all entries $\{a_{rp}^{(i,j)} \mid i,j = 1, \ldots, d, \ r = 1, \ldots, N_i, p = 1, \ldots, N_j\}$ of the matrix $X_{N_1,\ldots,N_d}$ forms a Gaussian family which is determined by*

$$a_{rp}^{(i,j)} = \overline{a_{pr}^{(j,i)}} \qquad \text{for all } i,j = 1,\ldots,d, \ r = 1,\ldots,N_i, \ p = 1,\ldots,N_j$$

*and the prescription of mean zero and covariance*

$$E[a_{rp}^{(i,j)} a_{qs}^{(k,l)}] = \frac{1}{n} \delta_{rs} \delta_{pq} \cdot \sigma(i,j;k,l),$$

*where we put*

$$n := N_1 + \cdots N_d.$$

*Then, for $n \to \infty$ such that*

$$\lim_{n \to \infty} \frac{N_i}{n} = \alpha_i \qquad \text{for all } i = 1, \ldots, d,$$

*the matrix $X_{N_1,\ldots,N_d}$ has a limiting eigenvalue distribution whose Cauchy transform $G(z)$ is determined by*

$$G(z) = \mathrm{tr}_\alpha(\mathcal{G}(z)),$$

*where $\mathcal{G}(z)$ is an $M_\alpha(\mathbb{C})$-valued analytic function on the upper complex half plane, which is uniquely determined by the facts that (11) holds and that it satisfies for all $z$ in the upper complex half plane the matrix equation*

$$z\mathcal{G}(z) = I_d + \eta_\alpha(\mathcal{G}(z)) \cdot \mathcal{G}(z). \tag{13}$$

In principle, the statement for the rectangular situation can be reduced to the case of square blocks by cutting the rectangular blocks into smaller square blocks (at least asymptotically); thus Theorem 4 can be deduced from Theorem 2. However, we find it more instructive to adapt our proof

of Theorem 2 to the rectangular situation - this makes the structure of the weighted covariance mapping and the appearance of the weighted trace more transparent. We will present this proof of Theorem 4 also in the next section. Let us also mention that the considerations of Shlyakhtenko for Gaussian band matrices were extended to rectangular matrices by Benaych-Georges [20]; thus some of his results are closely related to the special case of band matrices in our Theorem 4.

## 3.3 Wishart type block matrices

Another type of block matrices appearing in applications is of a Wishart type. Here the block matrix $H$ itself is rectangular and one is interested in the eigenvalue distribution of $HH^*$. Let us write $H = (A^{(i,j)})_{\substack{i=1,\ldots,r \\ j=1,\ldots,s}}$, where each block $A^{(i,j)}$ is a Gaussian $M_i \times N_j$ random matrix. Put $M := M_1 + \cdots + M_r$ and $N := N_1 + \cdots N_s$, so $H$ is an $M \times N$-matrix. The calculation of the eigenvalue distribution of $HH^*$ can be reduced to the situation treated in the previous section by the following trick. Consider

$$X = \begin{bmatrix} 0 & H \\ H^* & 0 \end{bmatrix}. \tag{14}$$

With $n = M + N$, this is a selfadjoint $n \times n$-matrix and can be viewed as a $d \times d$-block matrix of the form considered in the previous section, where $n = M + N$ is decomposed into $d := r + s$ parts according to

$$M + N = M_1 + \cdots + M_r + N_1 + \cdots N_s.$$

Thus we can use Theorem 4 to get the eigenvalue distribution of $X$ if the $M_i$ and $N_j$ go to infinity with a limit for their ratio.

The only remaining question is how to relate the eigenvalues of $X$ with those of $HH^*$. This is actually quite simple, we only have to note that all the odd moments of $X$ are zero and

$$X^2 = \begin{bmatrix} HH^* & 0 \\ 0 & H^*H \end{bmatrix}$$

Thus the eigenvalues of $X^2$ are the eigenvalues of $HH^*$ together with the eigenvalues of $H^*H$. Furthermore, note that $HH^*$ is an $M \times M$ and $H^*H$ is an $N \times N$ matrix. Assuming that $M < N$ (otherwise exchange the role of

$H$ and $H^*$) we have then that the eigenvalues of $H^*H$ are the eigenvalues of $HH^*$ plus $N - M$ additional zeros. In terms of the unnormalized trace Tr this just means

$$\text{Tr}(X^{2k}) = \text{Tr}((HH^*)^k) + \text{Tr}((H^*H)^k) = 2\text{Tr}((HH^*)^k)$$

Rewritten in normalized traces this gives

$$\text{tr}_n(X^{2k}) = \frac{2M}{n}\text{tr}_M\left((HH^*)^k\right).$$

Putting all these together, it results in the following relation between the Cauchy transform $G_{X^2}$ of $X^2$ and the Cauchy transform $G_{HH^*}$ of $HH^*$:

$$G_{HH^*}(z) = \frac{M+N}{2M}G_{X^2}(z) - \frac{N-M}{2M}\frac{1}{z}. \tag{15}$$

Finally, we should also rewrite our equation for the Cauchy transform $G_X$ of $X$ in terms of $G_{X^2}$. Since $X$ is even, both are related by

$$z \cdot G_{X^2}(z^2) = G_X(z).$$

By noting that the operator-valued version $\mathcal{G}(z)$ depends only on $z^2$, we can introduce a quantity $\mathcal{H}$ by

$$z \cdot \mathcal{H}(z^2) = \mathcal{G}(z).$$

Then we have

$$\lim_{n\to\infty} G_{X^2}(z) = \text{tr}_\alpha[\mathcal{H}(z)], \tag{16}$$

and the equation (12) for $\mathcal{G}$ becomes

$$z\mathcal{H}(z) = I_d + z\eta\big(\mathcal{H}(z)\big) \cdot \mathcal{H}(z). \tag{17}$$

The equations (16), (17), together with (15) determine the asymptotic eigenvalue distribution of $HH^*$ by

$$\lim_{M+N\to\infty} G_{HH^*}(z) = \theta\text{tr}_\alpha\mathcal{H}(z) - \frac{\theta_0}{z}, \tag{18}$$

where

$$\theta = \lim_{N\to\infty} \frac{M+N}{2M} = \frac{1}{2\sum_{j=1}^r \alpha_j}, \quad \theta_0 = \lim_{N\to\infty} \frac{N-M}{2M} = \frac{\sum_{j=r+1}^d \alpha_j - \sum_{j=1}^r \alpha_j}{2\sum_{j=1}^r \alpha_j}.$$

13

Note that the subtraction of a pole at 0 in (15) reflects just the occurrence of spurious zeros in $X^2$ which are coming from $H^*H$.

It would seem more adequate to have access to the information about the eigenvalues of $HH^*$ without having to invoke the eigenvalue distribution of $H^*H$. To simplify notations we will restrict in the following to the situation where all blocks $A^{(i,j)}$ are square matrices i.e., $N_1 = \cdots = N_r = M_1 = \cdots = M_s$, and thus $N = rN_1$ and $M = sN_1$. (Note that $H$ itself might still be rectangular, i.e., we are allowing that $r \neq s$.) One can of course write our $(r+s) \times (r+s)$ matrix $\mathcal{H}$ as a $2 \times 2$-block matrix

$$\mathcal{H}(z) = \begin{bmatrix} \mathcal{G}_1(z) & \mathcal{G}_3(z) \\ \mathcal{G}_3^*(z) & \mathcal{G}_2(z) \end{bmatrix}$$

where $\mathcal{G}_1(z)$, $\mathcal{G}_2(z)$, and $\mathcal{G}_3(z)$ are $r \times r$-, $s \times s$-, and $r \times s$-matrices, respectively. From our considerations in the next section is fairly easy to see that

$$\lim_{n \to \infty} G_{HH^*}(z) = \text{tr}_r\big(\mathcal{G}_1(z)\big). \tag{19}$$

One should note that the quadratic matrix equation (17) for $\mathcal{H}$ will in general couple all the various $\mathcal{G}_i$, so that even in this formulation the calculation of the distribution of $HH^*$ will still involve $H^*H$. However, there are special situations where one can eliminate $H^*H$ quite easily from our equations. We will examine one such situation, which appears in most applications to wireless communications, in more detail in the rest of this section.

Let us consider the special version of our Wishart type block matrices where the blocks of $H$ themselves are non-selfadjoint Gaussian random matrices (which are still, as before, square matrices). In this case the covariance $\sigma$ for the blocks of (14) couples effectively only blocks from $H$ with blocks from $H^*$ and thus $\eta : M_{r+s}(\mathbb{C}) \to M_{r+d}(\mathbb{C})$ maps $M_r(\mathbb{C}) \oplus M_s(\mathbb{C})$ to itself by

$$\eta : \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix} \mapsto \begin{bmatrix} \eta_2(D_2) & 0 \\ 0 & \eta_1(D_1) \end{bmatrix},$$

where

$$\eta_1 : M_r(\mathbb{C}) \to M_s(\mathbb{C}) \qquad \text{and} \qquad \eta_2 : M_s(\mathbb{C}) \to M_r(\mathbb{C}).$$

One sees now easily that $\mathcal{H}$ must take on values in $M_r(\mathbb{C}) \oplus M_s(\mathbb{C})$, thus

$$\mathcal{H}(z) = \begin{bmatrix} \mathcal{G}_1(z) & 0 \\ 0 & \mathcal{G}_2(z) \end{bmatrix},$$

where $\mathcal{G}_1$ and $\mathcal{G}_2$ are $M_r(\mathbb{C})$-valued and $M_s(\mathbb{C})$-valued, respectively, analytic functions in the upper complex half plane. The equation (17) for $\mathcal{H}$ splits now into the two equations

$$z\mathcal{G}_1(z) = I_r + z\eta_2\big(\mathcal{G}_2(z)\big) \cdot \mathcal{G}_1(z)$$

and

$$z\mathcal{G}_2(z) = I_s + z\eta_1\big(\mathcal{G}_1(z)\big) \cdot \mathcal{G}_2(z).$$

One can eliminate $\mathcal{G}_2$ from those equations by solving the second equation for $\mathcal{G}_2$ and inserting this into the first equation, yielding

$$z\mathcal{G}_1(z) = I_r + \eta_2\Big(\big(I_s - \eta_1\big(\mathcal{G}_1(z)\big)\big)^{-1}\Big) \cdot \mathcal{G}_1(z)$$

Together with (19) this gives directly the eigenvalue distribution of $HH^*$.

# 4  Proof of the Main Theorems

## 4.1  Proof of Theorem 2

We will prove Theorem 2 by calculating the moments of the averaged eigenvalue distribution of the $n \times n$ random matrices $X_N$ in the limit $N \to \infty$. Note that the $m$-th moment of the averaged eigenvalue distribution is just given by the expectation of the normalized trace $\mathrm{tr}_n$ of the $m$-th power of $X_N$. For this we can calculate

$$E[\mathrm{tr}_n(X_N^m)] = \frac{1}{d} \sum_{i(1),\dots,i(m)=1}^{d} E\big[\mathrm{tr}_N\big(A^{(i(1),i(2))} \cdots A^{(i(m),i(1))}\big)\big]$$

$$= \frac{1}{dN} \sum_{i(1),\dots,i(m)=1}^{d} \sum_{r(1),\dots,r(m)=1}^{N} E\big[a_{r(1)r(2)}^{(i(1),i(2))} \cdots a_{r(m)r(1)}^{(i(m),i(1))}\big]$$

Now we need to invoke the Wick formula for calculating the expectation of a product of entries from our matrix. The Wick formula allows to express all higher moments in random variables from a Gaussian family in terms of the covariances by a nice combinatorial formula, by summing over all pairings of the appearing factors. In our case this gives

$$E\big[a_{r(1)r(2)}^{(i(1),i(2))} \cdots a_{r(m)r(1)}^{(i(m),i(1))}\big] = \sum_{\pi \in \mathcal{P}_2(m)} \prod_{(p,q) \in \pi} E[a_{r(p)r(p+1)}^{(i(p),i(p+1))} a_{r(q)r(q+1)}^{(i(q),i(q+1))}].$$

Here $\mathcal{P}_2(m)$ denotes the pairings of a set of $m$ elements. A pairing is just a decomposition into subsets (called blocks of the pairing) with two elements. The product in the above Wick formula runs over all blocks of the pairing $\pi$. If we now use the formula (10) for the covariances of the entries of our random matrix $X_N$, then we can continue with

$$E\left[a_{r(1)r(2)}^{(i(1),i(2))} \cdots a_{r(m)r(1)}^{(i(m),i(1))}\right]$$
$$= \sum_{\pi \in \mathcal{P}_2(m)} \prod_{(p,q) \in \pi} \frac{1}{n} \cdot \sigma\big(i(p), i(p+1); i(q), i(q+1)\big) \cdot \delta_{r(p)r(q+1)} \cdot \delta_{r(p+1)r(q)}.$$

In order to understand better the contribution of a pairing $\pi$ according to this formula, it will be useful to interpret this formula in the language of permutations. By $S_m$ we will denote the group of permutations on $m$ elements. Let us denote by $\gamma \in S_m$ the cyclic permutation $\gamma = (1, 2, \dots, m)$. If we identify a partition $\pi \in \mathcal{P}_2(m)$ with a permutation in $S_m$ – by declaring the blocks to cycles, i.e., $\pi(p) = q$ and $\pi(q) = p$ for $(p, q) \in \pi$ – then one recognizes easily that

$$\sum_{r(1),\dots,r(m)=1}^{N} \prod_{(p,q) \in \pi} \delta_{r(p)r(q+1)} \cdot \delta_{r(p+1)r(q)} = \sum_{r(1),\dots,r(m)=1}^{N} \prod_{p=1}^{m} \delta_{r(p)r(\gamma\pi(p))} = N^{\#\gamma\pi},$$

where $\#\gamma\pi$ denotes the number of cycles of the permutation $\gamma\pi$.

With this we can continue our above calculation of the expected trace of $X_N^m$ as follows:

$$E[\mathrm{tr}_n(X_N^m)] = \sum_{\pi \in \mathcal{P}_2(m)} n^{\#\gamma\pi - m/2 - 1} \frac{1}{d^{\#\gamma\pi}} \sum_{i(1),\dots,i(m)=1}^{d} \prod_{(p,q) \in \pi} \sigma\big(i(p), i(p+1); i(q), i(q+1)\big)$$

Note that for $d = 1$ (and $\sigma(1, 1; 1, 1) = 1$) we recover the well-known genus expansion

$$E[\mathrm{tr}_n(A^m)] = \sum_{\pi \in \mathcal{P}_2(m)} n^{\#\gamma\pi - m/2 - 1}$$

of a selfadjoint Gaussian $n \times n$ random matrix.

It is quite easy to see that

$$\#\gamma\pi - m/2 - 1 \leq 0 \qquad \text{for any } \pi \in \mathcal{P}_2(m),$$

16

and that we have equality there exactly for the so-called non-crossing pairings $\pi \in NC_2(m)$. The latter are those pairings $\pi$ of $1, 2, \ldots, 2m$ for which the blocks of $\pi$ can be drawn in such a way below the numbers that they do not cross. Another useful way of describing non-crossing pairings is the following recursive procedure: A pairing of $2m$ numbers is non-crossing if and only if one can find a block consisting of two consecutive numbers such that after removing this block one gets a non-crossing pairing of the $2m - 2$ remaining numbers. For illustration, here are the 5 of the 15 pairings of 6 elements which are non-crossing:
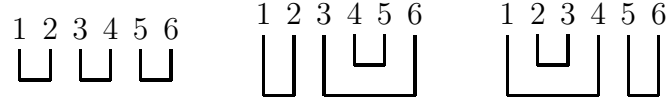
and
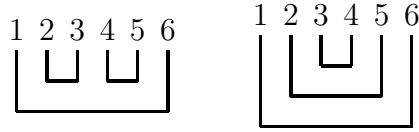
So in the limit, $N \to \infty$, only the non-crossing pairings survive in our formula and we get

$$\lim_{N \to \infty} E[\mathrm{tr}_n(X_N^m)] = \sum_{\pi \in NC_2(m)} \mathcal{K}_\pi,$$

where

$$\mathcal{K}_\pi := \frac{1}{d^{m/2+1}} \sum_{i(1),\ldots,i(m)=1}^{d} \prod_{(p,q) \in \pi} \sigma\big(i(p), i(p+1); i(q), i(q+1)\big)$$

In the case $d = 1$, all $\mathcal{K}_\pi$ would be equal and we would recover the moments of a semicircular distribution (in the form of the Catalan numbers counting non-crossing pairings). This would reproduce more or less the derivation of Wigner of his famous semicircle law. For $d > 1$, however, the $\mathcal{K}_\pi$ are in general different for different $\pi$, and, even worse, there does not exist a straightforward recursive way of expressing $\mathcal{K}_\pi$ in terms of "smaller" $\mathcal{K}_\sigma$. Thus we are outside the realm of the usual techniques of free probability theory.

However, one can save most of those techniques by going over to an "operator-valued" level. The main point of such an operator-valued approach

is to write $\mathcal{K}_\pi$ as the trace of a $d \times d$-matrix $\kappa_\pi$, and then realize that $\kappa_\pi$ has the usual nice recursive structure of semicircular elements.

Namely, let us define a matrix

$$\kappa_\pi = (\kappa_\pi(i,j))_{i,j=1}^d$$

by

$$\kappa_\pi(i,j) := \sum_{i(1)...,i(m),i(m+1)=1}^d \delta_{ii(1)} \delta_{ji(m+1)} \prod_{(p,q)\in\pi} \frac{1}{d} \cdot \sigma\big(i(p),i(p+1);i(q),i(q+1)\big).$$

Then clearly we have

$$\mathcal{K}_\pi = \mathrm{tr}_d(\kappa_\pi).$$

Furthermore, the value of $\kappa_\pi$ can be determined by an iterated application of our covariance mapping $\eta : M_d(\mathbb{C}) \to M_d(\mathbb{C})$. Namely, the value of $\kappa_\pi$ is given by an iterated application of this mapping $\eta$ according to the nesting of the blocks of $\pi$. If one identifies a non-crossing pairing with a putting of brackets, then the way that $\eta$ has to be iterated is quite obvious. Let us clarify these remarks with an example.
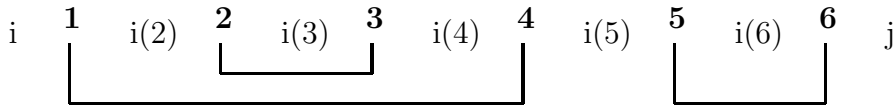
Consider the non-crossing pairing

$$\pi = \{(1,4),(2,3),(5,6)\} \in NC_2(6) \qquad \hat{=} \qquad \text{⊔⊔ ⊔}$$

The corresponding $\kappa_\pi$ is

$$\kappa_\pi(i,j) = \frac{1}{d^3} \sum_{i(2),i(3),i(4),i(5),i(6)=1}^d \sigma(i,i(2);i(4),i(5)) \cdot \sigma(i(2),i(3);i(3),i(4))$$
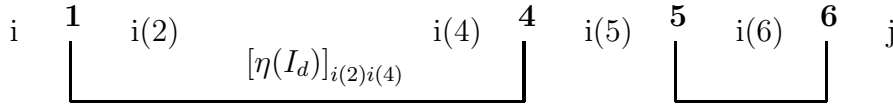
$$\cdot \, \sigma(i(5),i(6);i(6),j).$$

$\pi$ is a non-crossing pairing of the numbers $\mathbf{1,2,3,4,5,6}$. Let us put the indices $i = i(1), i(2), i(3), i(4), i(5), i(6), i(7) = j$ between those numbers. Each block of $\pi$ corresponds to one factor $\sigma$ in $\kappa_\pi$ and the arguments in this $\sigma$ are the indices to the left and to the right of the two numbers which are paired by this block.

i   **1**   i(2)   **2**   i(3)   **3**   i(4)   **4**   i(5)   **5**   i(6)   **6**   j

18

As mentioned before, for a non-crossing pairing one always has at least one block consisting of neighbours; in our case such a block is $(\mathbf{2}, \mathbf{3}) \in \pi$ (actually $(\mathbf{5}, \mathbf{6})$ is another one). This means that we can sum over the corresponding index $i(3)$ without interfering with other blocks, giving

$$\kappa_\pi(i,j) = \frac{1}{d^2} \sum_{i(2),i(4),i(5),i(6)=1}^{d} \sigma(i, i(2); i(4), i(5)) \cdot \sigma(i(5), i(6); i(6), j)$$

$$\cdot \frac{1}{d} \sum_{i(3)=1}^{d} \sigma(i(2), i(3); i(3), i(4))$$

$$= \frac{1}{d^2} \sum_{i(2),i(4),i(5),i(6)=1}^{d} \sigma(i, i(2); i(4), i(5)) \cdot \sigma(i(5), i(6); i(6), j) \cdot \eta(1)_{i(2)i(4)}$$
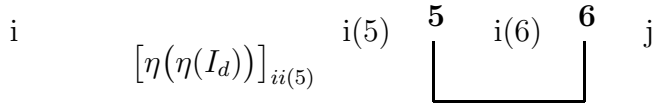
Effectively we have removed the block $(\mathbf{2}, \mathbf{3})$ and replaced it by the matrix $\eta(I_d)$; we are left with

i $\quad$ **1** $\quad$ i(2) $\quad\quad\quad\quad\quad\quad\quad$ i(4) $\quad$ **4** $\quad$ i(5) $\quad$ **5** $\quad$ i(6) $\quad$ **6** $\quad$ j

$$[\eta(I_d)]_{i(2)i(4)}$$

Now the block $(\mathbf{1}, \mathbf{4})$ of $\pi$ has become a block consisting of neighbours and we can do the summation over $i(2)$ and $i(4)$ without interfering with the other blocks, thus yielding

$$\kappa_\pi(i,j) = \frac{1}{d} \sum_{i(5),i(6)=1}^{d} \sigma(i(5), i(6); i(6), j)$$

$$\cdot \frac{1}{d} \sum_{i(2),i(4)=1}^{d} \sigma(i, i(2); i(4), i(5)) \cdot [\eta(I_d)]_{i(2)i(4)}$$

$$= \frac{1}{d} \sum_{i(5),i(6)=1}^{d} \sigma(i(5), i(6); i(6), j) \cdot \big[\eta\big(\eta(I_d)\big)\big]_{ii(5)}$$

We have now removed the block $(\mathbf{1}, \mathbf{4})$ and the effect of this was that we had to apply $\eta$ to whatever was embraced by this block (in our case, $\eta(I_d)$). We remain with

i $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ i(5) $\quad$ **5** $\quad$ i(6) $\quad$ **6** $\quad$ j

$$\big[\eta\big(\eta(I_d)\big)\big]_{ii(5)}$$

19

Now we can do the summation over $i(5)$ and $i(6)$ corresponding to the last block $(\mathbf{5}, \mathbf{6})$ of $\pi$, which results in

$$\kappa_\pi(i, j) = \sum_{i(5)=1}^{d} \left[\eta\big(\eta(I_d)\big)\right]_{ii(5)} \big(\frac{1}{d} \sum_{i(6)=1}^{d} \sigma(i(5), i(6); i(6), j)\big)$$

$$= \sum_{i(5)=1}^{d} \left[\eta\big(\eta(I_d)\big)\right]_{ii(5)} \cdot \eta(1)_{i(5)j}$$

$$= \left[\eta\big(\eta(I_d)\big) \cdot \eta(I_d)\right]_{ij}$$

Thus we finally have

$$\kappa_\pi = \eta\big(\eta(I_d)\big) \cdot \eta(I_d),$$

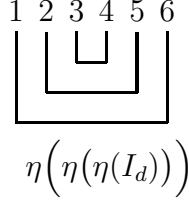which corresponds to the bracket expression

$$(X(XX)X)(XX).$$

In the same way every non-crossing pairing results in an iterated application of the mapping $\eta$. For the five non-crossing pairings of six elements one gets the following results:



$$\eta(I_d) \cdot \eta(I_d) \cdot \eta(I_d)$$



$$\eta(I_d) \cdot \eta\big(\eta(I_d)\big)$$



$$\eta\big(\eta(I_d)\big) \cdot \eta(I_d)$$



$$\eta\big(\eta(I_d) \cdot \eta(I_d)\big)$$

and

$$\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

$$\eta\Big(\eta\big(\eta(I_d)\big)\Big)$$

Thus for $m = 6$ we get

$$\lim_{n \to \infty} E[\mathrm{tr}(A^6)]$$
$$= \mathrm{tr}_d\Big\{\eta(I_d)\cdot\eta(I_d)\cdot\eta(I_d)+\eta(I_d)\cdot\eta\big(\eta(I_d)\big)+\eta\big(\eta(I_d)\big)\cdot\eta(I_d)+\eta\big(\eta(I_d)\cdot\eta(I_d)\big)+\eta\big(\eta\big(\eta(I_d)\big)\big)\Big\}.$$

Let us summarize our calculations for general moments: We have

$$\lim_{N \to \infty} E[\mathrm{tr}_n(X_N^m)] = \mathrm{tr}_d\Big\{\sum_{\pi \in NC_2(m)} \kappa_\pi\Big\},$$

where $\kappa_\pi$ are $d \times d$ matrices, determined in a recursive way as above, given by an iterated application of the mapping $\eta$.

Let us denote by $\mathcal{E}$ the limiting operator-valued moments of $X_N$, i.e.,

$$\mathcal{E}(X^m) := \sum_{\pi \in NC_2(m)} \kappa_\pi,$$

so that we have

$$\lim_{N \to \infty} E[\mathrm{tr}_n(X_N^m)] = \mathrm{tr}_d(\mathcal{E}(X^m)).$$

An element $X$ whose operator-valued moments $\mathcal{E}(X^m)$ are calculated in such a way is called an *operator-valued semicircular element.*

The above description of the moments $\mathcal{E}(X^m)$ can, similarly as for an ordinary semicircular variable, be reformulated in a recursive way which leads directly to our equation (12). To get this recursive description, one has to observe that if $\pi$ is a non-crossing pairing of $m$ elements, and $(1, r)$ is the block of $\pi$ containing 1, then the remaining blocks of $\pi$ must fall into two classes, those making up a non-crossing pairing of the numbers $2, 3, \ldots, r - 1$ and those making up a non-crossing pairing of the numbers $r + 1, r + 2, \ldots, m$. Let us call the former pairing $\pi_1$ and the latter $\pi_2$, so that we can write $\pi = (1, r) \cup \pi_1 \cup \pi_2$. Then the above description of $k_\pi$ shows that

$$\kappa_\pi = \eta\big(\kappa_{\pi_1}\big) \cdot \kappa_{\pi_2}.$$

21

This results for the operator valued moments in the recurrence relation

$$\mathcal{E}[X^m] = \sum_{k=0}^{m-2} \eta\big(\mathcal{E}[X^k]\big) \cdot \mathcal{E}[X^{m-k-2}].$$

If we go over to the corresponding generating power series,

$$\mathcal{M}(z) = \sum_{m=0}^{\infty} \mathcal{E}[X^m] z^m,$$

then this yields the relation

$$\mathcal{M}(z) = I_d + z^2 \eta\big(\mathcal{M}(z)\big) \cdot \mathcal{M}(z).$$

Note that $M(z) := \mathrm{tr}_d(\mathcal{M}(z))$ is the generating power series of the moments $E[\mathrm{tr}_N(X_N^m)]$ for $N \to \infty$, thus it is preferable to go over from $\mathcal{M}(z)$ to the corresponding operator-valued Cauchy transform

$$\mathcal{G}(z) := \frac{1}{z}\mathcal{M}(1/z).$$

For this the above equation takes on the form

$$z\mathcal{G}(z) = I_d + \eta(\mathcal{G}(z)) \cdot \mathcal{G}(z),$$

which is exactly our equation (12). Furthermore, we have for the Cauchy transform $G$ of the limit eigenvalue distribution of our block matrix $X_N$ that

$$G(z) = \frac{M(1/z)}{z} = \mathrm{tr}_d(\frac{\mathcal{M}(1/z)}{z}) = \mathrm{tr}_d(\mathcal{G}(z)).$$

Let us finally make a few remarks about the support of the limiting eigenvalue distribution and the validity of equation (12) for all z in the upper half plane.

First note that the number of non-crossing pairings of $2k$ elements is given by the Catalan numbers which behave asymptotically for large $k$ like $4^k$. (Of course, for $m$ odd there are no pairings of $m$ elements at all.) This implies that we can estimate the (operator) norm of the matrix $\mathcal{E}(X^m)$ by

$$\|\mathcal{E}(X^m)\| \leq \|\eta\|^m \cdot \#NC_2(m) \leq c \cdot \|\eta\|^m \cdot 2^m,$$

where $c$ is a constant independent of $m$. Applying $\text{tr}_d$, this yields that the support of the limiting eigenvalue distribution of $X_N$ is contained in the interval $[-2\|\eta\|, +2\|\eta\|]$. Clearly, since all odd moments are zero, the measure is symmetric. Furthermore, the above estimate on the operator-valued moments $\mathcal{E}(X^m)$ shows that

$$\mathcal{G}(z) = \sum_{k=0}^{\infty} \frac{\mathcal{E}(X^{2k})}{z^{2k+1}}$$

is a power series expansion in $1/z$ of $\mathcal{G}(z)$, which converges in a neighborhood of $\infty$. Since the mapping

$$D \mapsto \frac{1}{z}I_d + \frac{1}{z}\eta(D) \cdot D$$

is a contraction for $|z|$ sufficiently large, $\mathcal{G}(z)$ is, for large $z$, uniquely determined as the solution of the equation (12).

In order to realize that $\mathcal{G}(z)$ can actually be extended to an analytic matrix-valued function on the whole upper complex half plane one has to note that the equation

$$\lim_{n \to \infty} E[\text{tr}_n(X_N^m)]$$

$$= \sum_{\pi \in NC_2(m)} \frac{1}{d^{m/2+1}} \sum_{i(1),\dots,i(m)=1}^{d} \prod_{(p,q) \in \pi} \sigma\big(i(p), i(p+1); i(q), i(q+1)\big)$$

can also be read in the way that the limiting moments of $X_N$ are given by the moments of a $d \times d$-matrix

$$X = (c_{ij})_{i,j=1}^d,$$

where the entries of $X$ form a circular/semicircular family living in some non-commutative probability space $(\mathcal{A}, \varphi)$, with covariance $\varphi(c_{ij}c_{kl}) = \sigma(i, j; k, l)$:

$$\lim_{n \to \infty} E[\text{tr}_n(X_N^m)] = \varphi \otimes \text{tr}_d(X^m).$$

What we showed above was then essentially that $X$ is an operator-valued semicircular variable over $M_d(\mathbb{C})$, where the expectation

$$\mathcal{E} : M_d(\mathcal{A}) = M_d(\mathbb{C}) \otimes \mathcal{A} \to M_d(\mathbb{C})$$

23

is given by applying $\varphi$ entrywise, i.e., $\mathcal{E} = 1 \otimes \varphi$. In this language,

$$\mathcal{G}(z) = \mathcal{E}\left(\frac{1}{z - X}\right),$$

which shows that it is actually an analytic function on the whole upper complex half plane. The validity of (12) for all $z$ in the upper half plane follows then by analytic continuation.

## 4.2   Proof of Theorem 4

The proof of Theorem 4 is very similar to the one of Theorem 2. We will concentrate here mainly on the modifications compared to the latter one.

Again, one calculates the averaged traces of powers by invoking the Wick formula. This yields

$$E[\mathrm{tr}_n(X^m_{N_1,\dots,N_d})] = \frac{1}{n} \sum_{i(1),\dots,i(m)=1}^{d} \sum_{r(1)=1}^{N_{i(1)}} \cdots \sum_{r(m)=1}^{N_{i(m)}} \sum_{\pi \in \mathcal{P}_2(m)} \prod_{(p,q)\in\pi} \frac{1}{n}$$
$$\cdot \sigma\big(i(p), i(p+1); i(q), i(q+1)\big) \cdot \delta_{r(p)r(q+1)} \cdot \delta_{r(p+1)r(q)}$$
$$= \sum_{\pi \in \mathcal{P}_2(m)} n^{\#\gamma\pi - m/2 - 1} \sum_{i(1),\dots,i(m)=1}^{d} \prod_{c\in\gamma\pi} \frac{N_{i(c)}}{n} \prod_{(p,q)\in\pi} \sigma\big(i(p), i(p+1); i(q), i(q+1)\big).$$

Here $N_{i(c)}$ denotes the value $N_{i(k)}$ which belongs to the cycle $c$ of $\gamma\pi$. Note that the appearance of the $\sigma$-factors ensures that, although $i(k)$ might be different for different $k$ in the cycle $c$, all $N_{i(k)}$ for $k \in c$ are actually the same.

Thus in the limit $n \to \infty$, again only the non-crossing pairings survive and we get as before

$$\lim_{n\to\infty} E[\mathrm{tr}(X^m_{N_1,\dots,N_d})] = \sum_{\pi \in NC_2(m)} \mathcal{K}_\pi,$$

where now the contribution of a $\pi \in NC_2(m)$ is given by

$$\mathcal{K}_\pi := \sum_{i(1),\dots,i(m)=1}^{d} \prod_{c\in\gamma\pi} \alpha_{i(c)} \prod_{(p,q)\in\pi} \sigma\big(i(p), i(p+1); i(q), i(q+1)\big).$$

24

As before we can write this as a trace of a matrix-valued object – but this time everything is weighted with the vector $\alpha$. Namely, let us define the matrix

$$\kappa_\pi = (\kappa_\pi(i,j))_{i,j=1}^d$$

by

$$\kappa_\pi(i,j) := \sum_{i(1)...,i(m),i(m+1)=1}^d \delta_{ii(1)}\delta_{ji(m+1)}$$
$$\prod_{1\notin c\in\gamma\pi} \alpha_{i(c)} \prod_{(p,q)\in\pi} \sigma\big(i(p),i(p+1);i(q),i(q+1)\big),$$

where this time we only take the product over those cycles of $\gamma\pi$ which do not contain the element 1. The factor $\alpha_{i(1)}$ corresponding to the cycle containing 1 will be used as a weighting factor for the remaining trace:

$$\mathcal{K}_\pi = \mathrm{tr}_\alpha(\kappa_\pi).$$

One can now check easily that the value of $\kappa_\pi$ is calculated in the same way as before, by an iterated application of the (now weighted) covariance mapping $\eta_\alpha$, nested according to the nesting of the blocks of the non-crossing pairing $\pi$.

Let us summarize our calculations: We have

$$\lim_{n\to\infty} E[\mathrm{tr}_n(X_{N_1,...,N_d}^m)] = \mathrm{tr}_\alpha\Big\{ \sum_{\pi\in NC_2(m)} \kappa_\pi \Big\},$$

where $\kappa_\pi$ are $d\times d$ matrices in $M_d^\alpha(\mathbb{C})$, determined in a recursive way as above, and given by an iterated application of the mapping $\eta_\alpha$.

The derivation of the equation (13) is then as before.

## 5   Results and Discussion

Our theorems give us the Cauchy transform $G$ of the asymptotic eigenvalue distribution of the considered block matrices in the form $G(z) = \mathrm{tr}_\alpha(\mathcal{G}(z))$, where $\mathcal{G}(z)$ is a solution to the matrix equation (12), (13), or (17). We recover the corresponding eigenvalue distribution $\mu$ from $G$ in the usual way, by invoking Stieltjes inversion formula

$$d\mu(x) = -\frac{1}{\pi} \lim_{\varepsilon\searrow 0} \Im G(x + i\varepsilon)dx, \tag{20}$$

25

where the limit is weak convergence of measures (as we will remark in Sec. 5.4).

Usually, there is no explicit solution for our matrix equation, so that we have to rely on numerical methods for solving those. Note that we do not get directly an equation for $G$. We first have to solve the matrix equation, then take the (weighted) trace of this solution. Thus, in terms of the entries of our matrix $\mathcal{G}$, we face a system of quadratic equations which we solve numerically using Newton's algorithm [21].

In many concrete cases the matrix $\mathcal{G}$ exhibits some special structure (i.e., some of its entries might be zero or some of them agree); taking this into account from the beginning reduces the complexity of the considered system of equations considerably. This special pattern of the entries of $\mathcal{G}$ depends of course on the considered block structure of our problem, i.e. on the covariance function $\eta$. Let us point out how one can determine this pattern of $\mathcal{G}$, relying on the knowledge of $\eta$. Let us concentrate on the equation (12) for square matrices. As we remarked in the previous section, for large $z$, one can produce a solution to this equation by iterating the map

$$D \mapsto \frac{1}{z}I_d + \frac{1}{z}\eta(D) \cdot D.$$

Since a pattern in the entries of $\mathcal{G}$ which is valid for large $z$ must, by analytic continuation, remain valid everywhere, we can use the above map to determine this pattern. Since the factor $1/z$ plays no role for recognizing such a pattern, we start, for example, with the identity matrix $D = I_d$ and iterate the mapping $D \mapsto I_d + \eta(D) \cdot D$ a few times until the pattern in the entries of $D$ stabilizes. This is then the wanted pattern for the entries of $\mathcal{G}$.

In this section we will consider concrete examples of block matrices and compare our solution from Theorem 2 or 4 with the result of simulations. First, the well-known Marchenko-Pastur law is derived analytically. Then the eigenvalue distribution of two different block matrix scenarios is calculated. We conclude this section with remarks on convergence and the numerics of the proposed method.

## 5.1   Marchenko-Pastur law

This is a warm-up example in which we re-derive the celebrated Marchenko-Pastur law from Theorem 4.

### 5.1.1 Proposition: Marchenko-Pastur

If $H_N$ is and $M \times N$ matrix of i.i.d. random variables of unit variance, then as $N \to \infty$, $N/M \to \lambda > 0$, the empirical law of the eigenvalues of $HH^*/M$ converges weakly with probability one to

$$\mu(dx) = (1-\lambda)^+ \delta_0 + \frac{\sqrt{(x-1-\lambda)^2 - 4\lambda}}{2\pi x} 1_{(1-\sqrt{\lambda})^2 \le x \le (1+\sqrt{\lambda})^2} dx \qquad (21)$$

*Proof.* Since this is a well known result, we just show how the final formula follows from Theorem 4 applied to $X$ given by (14) with $H$ replaced by $H/\sqrt{M+N}$, $d = 2$, $\alpha_1 = \lim_{M \to \infty} M/(M+N) = 1/(1+\lambda)$, $\alpha_2 = \lambda/(1+\lambda)$; we only consider the case $\lambda \ge 1$. From (15) we have

$$\lim_{N \to \infty} G_{HH^*/(M+N)}(z) = \frac{1+\lambda}{2} \mathrm{tr}_\alpha \mathcal{G}(z) - \frac{\lambda-1}{2z} = f(z),$$

where $\mathcal{G}(z) = \mathrm{diag}(f(z), g(z))$ satisfies equation (17) with $\eta(\mathrm{diag}(a,b)) = \mathrm{diag}(\alpha_2 b, \alpha_1 a)$; we used (19) to identify the answer with $f$. Thus

$$zf = 1 + \lambda z fg/(1+\lambda), \ zg = 1 + zfg/(1+\lambda).$$

After a calculation which includes renormalization (dilation by $1/(1+\lambda)$), this gives

$$\lim_{M \to \infty} G_{HH^*/M}(z) = \frac{f(z/(1+\lambda))}{1+\lambda} = \frac{z + (1-\lambda) - \sqrt{(z-1-\lambda)^2 - 4\lambda}}{2z},$$

which is the Cauchy transform of the Marchenko-Pastur law, see for example [22, page 102] $\qquad \square$

## 5.2 Block Toeplitz matrices

In this section we show how to use Theorem 2 to identify the limit laws of block Toeplitz matrices that were considered in [12]. The interesting feature of this example is that $\mathcal{G}$ is non-diagonal. In this example we present only details for the case $d = 3$; in the figures we also show our result and comparison with simulation results for the cases $d = 4$ and $d = 5$.

Suppose $A, B, C$ are independent selfadjoint $N \times N$ matrices with i.i.d. entries of unit variance, and consider the block Toeplitz matrix:

$$X = \frac{1}{\sqrt{3N}} \begin{bmatrix} A & B & C \\ B & A & B \\ C & B & A \end{bmatrix}.$$

Then as $N \to \infty$ the spectral law of $X$ converges almost surely to a deterministic law, see [12]. Since the limit does not depend on the distribution of the entries, we can assume that they are Gaussian and apply Theorem 2. This determines the Cauchy transform of the limit law as the normalized trace of $\mathcal{G}$; the latter is determined by Equation (12). For $\mathcal{G}$ of the form

$$
\mathcal{G} = \begin{bmatrix} f & 0 & h \\ 0 & g & 0 \\ h & 0 & f \end{bmatrix},
$$

$\eta$ acts as follows:

$$
\eta(\mathcal{G}) = \frac{1}{3} \begin{bmatrix} 2\,f + g & 0 & g + 2\,h \\ 0 & 2\,f + g + 2\,h & 0 \\ g + 2\,h & 0 & 2\,f + g \end{bmatrix}.
$$

Since the above pattern of $\mathcal{G}$ is preserved under the mapping $D \mapsto I_3 + \eta(D) \cdot D$, we know that our wanted solution must have this form. So (12) gives the following system of equations:

$$
zf = 1 + \frac{g\,(f + h) + 2\,(f^2 + h^2)}{3}, \tag{22}
$$

$$
zg = 1 + \frac{g\,(g + 2\,(f + h))}{3}, \tag{23}
$$

$$
zh = \frac{4\,f\,h + g\,(f + h)}{3}. \tag{24}
$$

This system of equations simplifies to two quadratic equations which can be solved by symbolic software; the density is a mixture of the semicircle law and another curve, but the exact solutions were not helpful in the selection of the appropriate root. Fig. 2(a) illustrates the histogram and the theoretical density obtained numerically from applying the Stieltjes inversion formula on the Cauchy transform $G(z) = (2f(z) + g(z))/3$, $\mu(t) = -\frac{1}{\pi} \lim_{s \to 0} \Im G(t + is)$. The numerical solution is calculated by solving the resulting system of equations using Newton's method. The histograms are calculated for 100 matrices with Gaussian blocks of size $100 \times 100$.

In Figures 2(b) and 2(c) we compare the simulation and the result of our method for the cases $d = 4$ with

$$
X = \frac{1}{\sqrt{4N}} \begin{bmatrix} A & B & C & D \\ B & A & B & C \\ C & B & A & B \\ D & C & B & A \end{bmatrix},
$$

and $d = 5$ with

$$X = \frac{1}{\sqrt{5N}} \begin{bmatrix} A & B & C & D & E \\ B & A & B & C & D \\ C & B & A & B & C \\ D & C & B & A & B \\ E & D & C & B & A \end{bmatrix}.$$

It might be of interest to point out that the double limit, first as block sizes $N \to \infty$ and then $d \to \infty$ of block Toeplitz matrices gives the semicircle law. This can be seen by passing to $d \times d$ Toeplitz matrices with semicircular entries. Since the combinatorial estimates from [23] apply to this case, the moments are given as sums over pair partitions. For semicircular elements, only non-crossing partitions contribute, and each such partitions contributes 1.

## 5.3 Wishart matrix of rectangular block Toeplitz matrices

In this section the introduced problem of calculating the spectrum for a MIMO channel is addressed, where it is of interest to calculate the distribution of the Wishart matrix of rectangular block Toeplitz matrices. First for the case $n_T = n_R$, a proposition is presented that simplifies the process to calculate the desired spectrum. In the Example subsection, first the spectrum for a MIMO system where $n_T = n_R$ is calculated. Then the problem for the $n_R = 2n_T$ is tackled and the results are presented.

Assuming $n_T = n_R = N$ and denoting the channel matrix with $H_N$, the following proposition draws the way to calculate the eigenvalue distribution of $H_N^* H_N$.

### 5.3.1 Proposition

As $N \to \infty$, $d = 2K + L - 1$, $n = Nd$ the spectral law of $H_N H_N^*/n$ converges weakly with probability one to the deterministic probability measure which is a mixture of $K$ densities with Cauchy-Stieltjes transform

$$\lim_{N \to \infty} G_{HH^*/n}(z) = \frac{1}{d} \sum_{j=1}^{K} f_j(z).$$

Functions $f_j$ are each a Cauchy transform of a probability measure and the following conditions hold.

1. $f_j = f_{K+1-j}$ for $1 \leq j \leq K$

2. There exist Cauchy transforms $g_1, g_2, \ldots, g_{K+L-1}$ of some probability measures such that $g_j = g_{K+L-j}$ for $1 \leq j \leq K + L - 1$ with the property that

$$\mathcal{G}(z) = \operatorname{diag}(f_1, f_2, \ldots, f_K, g_1, g_2, \ldots, g_{K+L-1})$$

satisfies (17) with

$$\eta(\mathcal{G})_{j,j} = \begin{cases} \frac{1}{d} \sum_{k=0}^{L-1} g_{j+k}, & \text{if } 1 \leq j \leq K, \\ \frac{1}{d} \sum_{k=1}^{j-K} f_k & \text{if } K + 1 \leq j \leq K + \min\{K, L\}, \\ \frac{1}{d} \sum_{k=j-K-L+1}^{j-K} f_k & \text{if } K + L < j \leq 2K \text{ and } L < K, \\ \frac{1}{d} \sum_{k=1}^{K} f_k & \text{if } 2K < j \leq K + L \text{ and } L \geq K, \\ \frac{1}{d} \sum_{k=j-K-L+1}^{K} f_k & \text{if } K + \max\{K, L\} < j \leq 2K + L - 1. \end{cases}$$

*Proof.* We omit the proof of almost sure convergence, see Section 5.4. To verify that $\mathcal{G}(z)$ is diagonal and its diagonal entries satisfy the symmetry condition we compute the diagonal entries of $\eta(D)$. For $1 \leq j \leq K$ we have

$$[\eta(D)]_{j,j} = \frac{1}{d} \sum_{k=0}^{L-1} d_{K+j+k, K+j+k}.$$

For $1 \leq j \leq \min\{K, L\}$ we have

$$[\eta(D)]_{K+j, K+j} = \frac{1}{d} \sum_{k=1}^{j} d_{k,k}.$$

For $\min\{K, L\} \leq j \leq \max\{K, L\}$ we have

$$[\eta(D)]_{K+j, K+j} = \begin{cases} \frac{1}{d} \sum_{k=j-L+1}^{j} d_{k,k} & L < K \\ \frac{1}{d} \sum_{k=1}^{K} d_{k,k} & L \geq K. \end{cases}$$

For $\max\{K, L\} \leq j \leq K + L - 1$ we have

$$[\eta(D)]_{K+j, K+j} = \frac{1}{d} \sum_{k=j-L+1}^{K} d_{k,k}.$$

We note that the symmetry conditions $f_j = f_{K+1-j}$ and $g_j = g_{K+L-j}$ are preserved under the mapping $D \mapsto I_d + \eta(D) \cdot D$. Since $\eta$ maps diagonal matrices into diagonal matrices, our solution $\mathcal{G}$ must be diagonal and the diagonal entries must satisfy the symmetry conditions as claimed. $\square$

30

### 5.3.2 Examples

In this part, first the developed theorems are used toward calculation of spectrum of a MIMO system with ISI ($L = 4$) and frame size of 4 ($K = 4$) while $n_T = n_R$. Then, for the simpler case of $K = 2$ and $L = 2$ when $n_R = 2n_T$ (or $n_T = 2n_R$) the spectrum is calculated.

Consider

$$H_N = \begin{bmatrix} A & B & C & D & 0 & 0 & 0 \\ 0 & A & B & C & D & 0 & 0 \\ 0 & 0 & A & B & C & D & 0 \\ 0 & 0 & 0 & A & B & C & D \end{bmatrix},$$

where $A, B, C, D$ are independent non-selfadjoint Gaussian $N \times N$-random matrices. It is also assumed that the impulse response of the channel from any transmitter antenna to any receiver antenna is identical and equal to $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. In this case $K = L = 4$, $d = 11$,

$$\mathcal{G}(z) = \mathrm{diag}(f_1, f_2, f_2, f_1, g_1, g_2, g_3, g_4, g_3, g_2, g_1),$$

and (17) yields the following system of equations

$$\begin{align}
zf_1 &= 1 + zf_1(g_1 + g_2 + g_3 + g_4)/11 \tag{25} \\
zf_2 &= 1 + zf_2(g_2 + 2g_3 + g_4)/11 \tag{26} \\
zg_1 &= 1 + zf_1 g_1/11 \tag{27} \\
zg_2 &= 1 + z(f_1 + f_2)g_2/11 \tag{28} \\
zg_3 &= 1 + z(f_1 + 2f_2)g_3/11 \tag{29} \\
zg_4 &= 1 + 2z(f_1 + f_2)g_4/11. \tag{30}
\end{align}$$

The limiting Cauchy transform is $G_{HH^*}(z) = (f_1 + f_2)/2$. After eliminating $g_j$ from the equations we get

$$\begin{align}
z &= \frac{1}{f_1} + \frac{1}{11 - f_1} + \frac{1}{11 - f_1 - f_2} + \frac{1}{11 - f_1 - 2f_2} + \frac{1}{11 - 2f_1 - 2f_2}, \\
z &= \frac{1}{f_2} + \frac{1}{11 - f_1 - f_2} + \frac{2}{11 - f_1 - 2f_2} + \frac{1}{11 - 2f_1 - 2f_2}.
\end{align}$$

Mathematica failed to solve this system in a reasonable time. Fortunately, the Newton's algorithm can be applied to the equations (25)-(30) and its answer matches the simulations (Fig. 3).

31

On the other hand, the channel matrix for a MIMO system with $n_R = 2n_T$ and frame length of $K = 2$ over a channel with ISI ($L = 2$) is as follows:

$$H_N = \begin{bmatrix} A & B & 0 \\ 0 & A & B \end{bmatrix},$$

where $A$ and $B$ are independent non-selfadjoint $2N \times N$ Gaussian random matrices. While one may calculate the spectrum of $HH^*$ using Theorem 4, it is also possible to consider

$$H_N = \begin{bmatrix} A & B & 0 \\ C & D & 0 \\ 0 & A & B \\ 0 & C & D \end{bmatrix}, \tag{31}$$

for $A$, $B$, $C$ and $D$ independent non-selfadjoint $N \times N$ Gaussian random matrices and calculate the spectrum of $HH^*$ using Theorem 2. The spectrum of $HH^*$ is calculated using the Newton's method which is depicted in Fig. 4.

## 5.4 General remarks on convergence

In Theorems 2 and 4, the convergence means weak convergence of the expected values of the empirical laws of the eigenvalues, which is sufficient to identify the limit. In applications, the appropriate mode of convergence is weak convergence of random empirical measures with probability one. Such convergence means that for a "typical" realization of a random matrix, the limiting law as determined from Theorems 2 and 4 is a good approximation when the size of the random matrix is large enough.

Results of this type are readily available for symmetric independent blocks with i.i.d. entries, see [12]. In the Gaussian case one can easily extend such results to non-Hermitian and dependent square blocks by noting the following:

- If $A$ consists of i.i.d. Gaussian random variables, then $A = A'/\sqrt{2} + A''/(\sqrt{2}i)$ where $A' = (A + A^*)/\sqrt{2}$ and $A'' = i(A - A^*)/\sqrt{2}$ are independent, Hermitian, and their off-diagonal entries are i.i.d Gaussian with the same variance as the entries of $A$.

- If $A, B$ are jointly Gaussian with respective entries that share the same correlation $\sigma$ (and are otherwise i.i.d), then $B = A + \sqrt{1 - \sigma^2}B'$, where

32

$B' = (A - \sigma B)/\sqrt{1 - \sigma^2}$ is independent of $A$ and has i.i.d. entries of the same variance as the entries of $A$.

For a Hermitian block matrix $X$ consisting of square blocks, these two transformations allow us to express $\mathrm{tr}_n(X^k)$ as a linear combination of the polynomials in independent Gaussian and Hermitian matrices. Under appropriate normalization (i.e. when the matrices are constructed from an infinite i.i.d family of unit variance, and the matrices are scaled by the square root of the dimension), the latter converge almost surely, see for example [24]. Since the limiting law is compactly supported, weak convergence of empirical measures holds with probability one.

## 5.5 General remarks on the numerics

The main inherent problem is that we have to find the "right" root. As we do not have a general automated recipe for doing this, we choose our solution by inspection. Theorem 2 guarantees uniqueness of an analytic solution under condition (11), but the validity of this condition is difficult to judge for a numerical solutions. It seems that in general there is only one solution which corresponds to a valid probability measure. Furthermore, as it is quite easily seen, the fact that $\mathcal{G}$ is essentially a generating power series in operator-valued moments implies that each diagonal entry of $\mathcal{G}$ is itself the Cauchy transform of some probability measure. Thus the solution we are looking for must exhibit the very strong property that all diagonal entries of $\mathcal{G}(z)$ yield a valid probability distribution when $z$ approaches the real line.

Our experience with this is reflected quite adequately in the following remark of Edelman and Rao [25] on the same kind of problem with their random matrix calculator: "Irrespective of whether the encoded probability measure is compactly supported or not, the $-1/z$ behaviour of the real part of Stieltjes transform (the principal value) as $z \to \pm\infty$ helps identify the correct root. In our experience, while multiple root curves might exhibit this behavior, invariably only one root will have an imaginary branch that, when normalized, will correspond to a valid probability measure. Why this always appears to be the case for the operational laws described is a bit of a mystery to us."

# 6    Conclusion

We use operator-valued free probability theory and develop a method to calculate the limiting spectra for block matrices. Using this method we compute the eigenvalue distribution for block-matrices arising in wireless communications. The agreement with simulation results is excellent. We expect our method to have a much wider applicability.

## Acknowledgement

We thank Ralf Müller for bringing the problem of block matrices in the context of MIMO to our attention.

# References

[1] G. J. Foschini and M. Gans, "On Limits of wireless communications in fading environment when using multi-element antennas," *Wireless Personal Communications*, Vol. 6, No. 3, pp. 331-335, 1998.

[2] S. Verdú, "Capacity region of Gaussian CDMA channels: The symbol synchronous case," *Proc. Allerton Conf. on Communication, control and computing*, Monticello, IL, pp.1025-1034, 1986.

[3] U. Madhow and M. Honi, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Communications*, Vol. 42, pp.3178-3188, Dec. 1994.

[4] S. Verdú, *Multiuser Detection*, Cambridge, UK: Cambridge University Press, 1998.

[5] E. G. Larsson and P. Stoica, *Space-time block coding for wireless communications*, Cambridge 2003.

[6] A. M. Tulino and S. Verdú, "Random matrices and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, 2004.

[7] D.N.C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Information Theory*, Vol. 46, No.1, pp. 171-188, Jan. 2000.

[8] A.L. Moustakas, S.H. Simon and A.M. Sengupta,"MIMO Capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Information Theory*, Vol. 49, No. 10, pp.2545-2561, Oct. 2003.

[9] R.R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels,"*IEEE Trans. Information Theory*, Vol.48, No.7, pp.2086-2091, July 2002.

[10] M. Debbah and R.R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Trans. Information Theory*, Vol.51, No.5, pp.1667-1690, May 2005.

[11] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, ser. CRM Monograph Series. Providence, RI: American Mathematical Society, 1992, vol. 1.

[12] T. Oraby, "The spectral laws of symmetric block-matrices with large random blocks," 2005, submitted.

[13] R.R. Müller, "A random matrix model of communication via antenna arrays,"*IEEE Trans. Information Theory*, Vol.48, No.9, pp.2495-2506, September 2002.

[14] D. Shlyakhtenko, "Random Gaussian band matrices and freeness with amalgamation," *Internat. Math. Res. Notices*, no. 20, pp. 1013–1025, 1996.

[15] D. Shlyakhtenko, "Gaussian random band matrices and operator-valued free probability theory," in *Quantum probability (Gdańsk, 1997)*, ser. Banach Center Publ. Warsaw: Polish Acad. Sci., 1998, vol. 43, pp. 359–368.

[16] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*, vol. 335 of London Mathematical Society Lecture Notes Series, Cambridge University Press, 2006.

[17] A. Nica, D. Shlyakhtenko, and R. Speicher, "Operator-valued distributions. I. Characterizations of freeness," *Int. Math. Res. Not.*, no. 29, pp. 1509–1538, 2002.

[18] R. Speicher, "Combinatorial theory of the free product with amalgamation and operator-valued free probability theory," *Mem. Amer. Math. Soc.*, vol. 132, no. 627, pp. x+88, 1998.

[19] D. Voiculescu, "Operations on certain non-commutative operator-valued random variables," *Astérisque*, no. 232, pp. 243–275, 1995, recent advances in operator algebras (Orléans, 1992).

[20] F. Benaych-Georges, "Rectangular random matrices, related free entropy and free Fisher's information," 2005, math.OA/0512081.

[21] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C*, 2nd ed. Cambridge: Cambridge University Press, 1992, the art of scientific computing.

[22] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, ser. Mathematical Surveys and Monographs. Providence, RI: American Mathematical Society, 2000, vol. 77.

[23] W. Bryc, A. Dembo, and T. Jiang, "Spectral measure of large random Hankel, Markov and Toeplitz matrices," *Ann. Probab.*, vol. 34, pp. 1–38, 2005, expanded version: arxiv.org/abs/math.PR/0307330.

[24] F. Hiai and D. Petz, "Asymptotic freeness almost everywhere for random matrices," *Acta Sci. Math. (Szeged)*, vol. 66, no. 3-4, pp. 809–834, 2000.

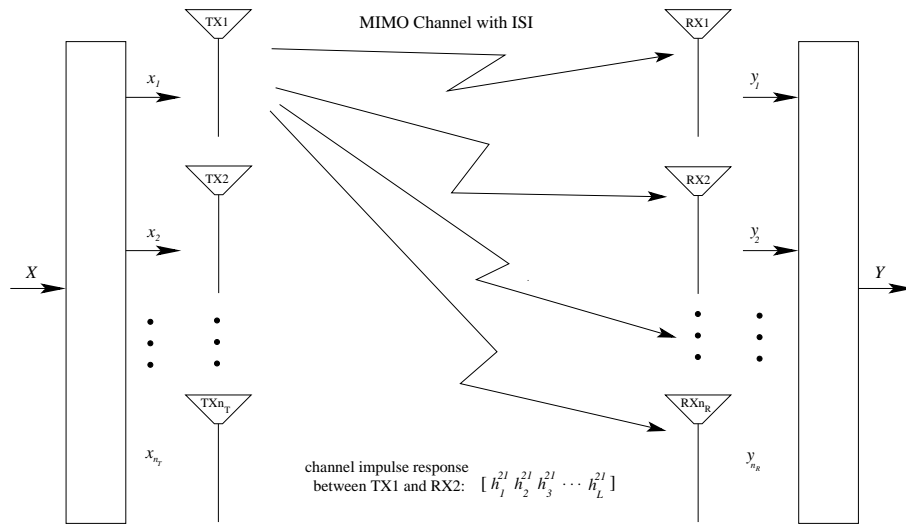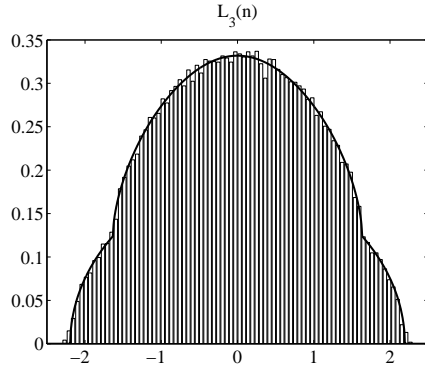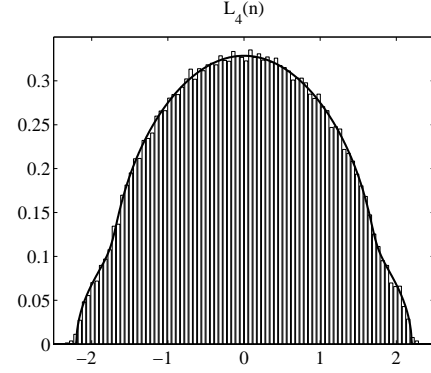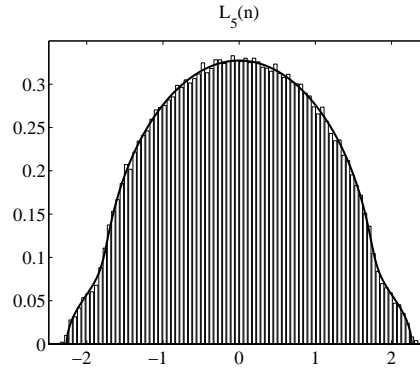[25] N. R. Rao and A. Edelman, "The polynomial method for random matrices, preprint, 2006.

Figure 1: Block diagram of a MIMO system with ISI.

(a) $3 \times 3$

(b) $4 \times 4$

(c) $5 \times 5$

Figure 2: Results for block Toeplitz matrices from Section 5.2. The histogram in each figure stands for the simulation result (100 realization of matrices with $100 \times 100$ Gaussian blocks) while the darker curve represents the numerical results of the proposed method.
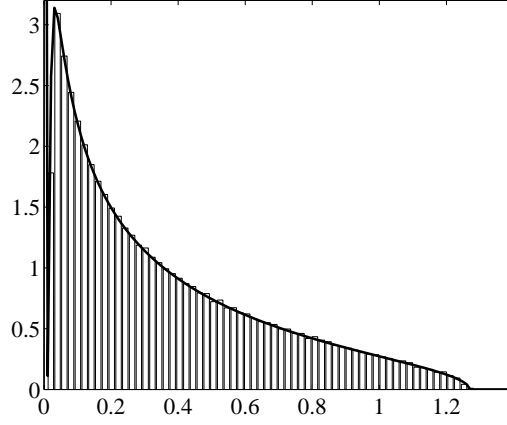
38

Figure 3: Superimposed theoretical density of the eigenvalues of complex normal $H_n H_n^*/n$ for a channel with ISI $L = 4$ and a MIMO system $n_R = n_T$ with frame length of $K = 4$ over its histogram for $N = 100$, based on 100 realizations.
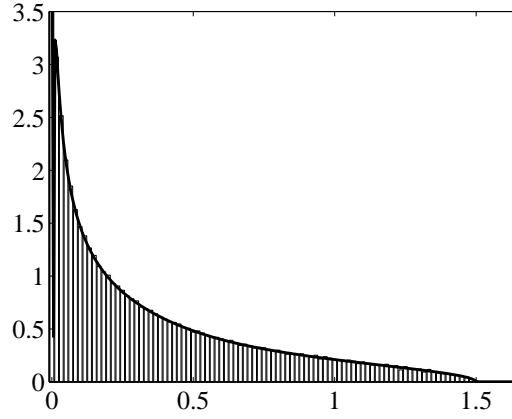


Figure 4: Superimposed theoretical density of the eigenvalues of complex normal $H_n H_n^*/n$ for a channel with ISI $L = 2$ and a MIMO system $n_R = 2n_T$ with frame length of $K = 2$ over its histogram for $N = 100$, based on 100 realizations.

# List of Figures